HBASE – HIVE INTEGRATION

By.Mr.Gopal Krishna
Sr.Hadoop Technical Architect
CCA 175-Spark and Hadoop Developer Certified Consultant

```
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
                        FROM HBASE SIDE
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
hbase(main):002:0> create 'hbasehive_emptable','empdetails'
0 row(s) in 1.9190 seconds

=> Hbase::Table - hbasehive_emptable
hbase(main):003:0>

hbase(main):003:0> scan 'hbasehive_emptable'
ROW                 COLUMN+CELL
0 row(s) in 0.1550 seconds

hbase(main):004:0>
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
                        FROM HIVE SIDE
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++


hive> create database hivehbasedb;
OK
Time taken: 0.493 seconds

hive> use hivehbasedb;
OK
Time taken: 0.02 seconds
```

ADD THE BELOW THREE JAR FILES – FOR HBASE –HIVE INTEGRATION

```
hive> add jar ${env:HIVE_HOME}/lib/zookeeper-3.4.6.jar;
Added [/home/gopalkrishna/INSTALL/apache-hive-1.2.1-bin/lib/zookeeper-3.4.6.jar] to class path
Added resources: [/home/gopalkrishna/INSTALL/apache-hive-1.2.1-bin/lib/zookeeper-3.4.6.jar]
hive> add jar ${env:HIVE_HOME}/lib/guava-14.0.1.jar;
Added [/home/gopalkrishna/INSTALL/apache-hive-1.2.1-bin/lib/guava-14.0.1.jar] to class path
Added resources: [/home/gopalkrishna/INSTALL/apache-hive-1.2.1-bin/lib/guava-14.0.1.jar]
hive> add jar ${env:HIVE_HOME}/lib/hive-hbase-handler-1.2.1.jar;
Added [/home/gopalkrishna/INSTALL/apache-hive-1.2.1-bin/lib/hive-hbase-handler-1.2.1.jar] to class
path
Added resources: [/home/gopalkrishna/INSTALL/apache-hive-1.2.1-bin/lib/hive-hbase-handler-1.2.1.jar]
hive>
```

 # CREATE A HIVE EXTERNAL TABLE WITH HBASE COLUMNS MAPPING
hive> CREATE EXTERNAL TABLE hive_emptable(key int,
firstname string,
lastname string,
yrsofexp int,
desg string)
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> WITH SERDEPROPERTIES
> (
> "hbase.columns.mapping"=
":key,empdetails:firstname,empdetails:lastname,empdetails:yrsofexp,empdetails:de
sg")
> TBLPROPERTIES ("hbase.table.name"="hbasehive_emptable");
OK
Time taken: 0.464 seconds
hive> describe formatted hive_emptable;
OK
 # col_name   data_type   comment

key   int   from deserializer
firstname   string   from deserializer
lastname   string   from deserializer
yrsofexp   int   from deserializer
desg   string   from deserializer

 # Detailed Table Information
Database: hivehbasedb
Owner: gopalkrishna
CreateTime: Thu Sep 14 00:46:48 PDT 2017
LastAccessTime: UNKNOWN
Protect Mode: None
Retention: 0
Location: hdfs://localhost:8020/user/hive/warehouse/hivehbasedb.db/hive_emptable
Table Type: EXTERNAL_TABLE
Table Parameters:
        COLUMN_STATS_ACCURATE   false
        EXTERNAL   TRUE
        hbase.table.name   hbasehive_emptable
        numFiles   0
        numRows   -1
        rawDataSize   -1
        storage_handler   org.apache.hadoop.hive.hbase.HBaseStorageHandler
        totalSize   0
        transient_lastDdlTime   1505375208

 # Storage Information
SerDe Library: org.apache.hadoop.hive.hbase.HBaseSerDe
InputFormat: null
OutputFormat: null
Compressed: No
Num Buckets: -1
Bucket Columns: []
Sort Columns: []
Storage Desc Params:

hbase.columns.mapping
:key,empdetails:firstname,empdetails:lastname,empdetails:yrsofexp,empdetails:desg
serialization.format   1
Time taken: 0.187 seconds, Fetched: 39 row(s)
hive>

hive> select * from hive_emptable;
OK
Time taken: 1.301 seconds
hive>

TO LOAD THE DATA INTO HIVE TABLE

hive> LOAD DATA LOCAL INPATH
'/home/gopalkrishna/PRAC/HBASE/HBase_Hive_InputData.log' INTO TABLE
hive_emptable;
FAILED: SemanticException [Error 10101]: A non-native table cannot be used as target for LOAD
hive>

hive> CREATE EXTERNAL TABLE initable(empid int,
>   firstname string,
>   lastname string,
>   yrsofexp int,
>   desg string)
>   ROW FORMAT DELIMITED
>   FIELDS TERMINATED BY '\t'
>   LINES TERMINATED BY '\n'
>   STORED AS TEXTFILE;
OK
Time taken: 0.105 seconds
hive> LOAD DATA LOCAL INPATH
'/home/gopalkrishna/PRAC/HBASE/HBase_Hive_InputData.log' INTO TABLE initable;
 Loading data to table hivehbasedb.initable
Table hivehbasedb.initable stats: [numFiles=2, numRows=0, totalSize=1142, rawDataSize=0]
OK
Time taken: 0.822 seconds

hive> SELECT * FROM initable;
OK

```
1000    RAVI    VARMA   9       TECHNOLOGY LEAD
1001    KOUSIK  MISTRA  11      ENTRYLEVEL ARCHITECT
1002    SAIED   FAISAL  21      DELIVERY HEAD
1003    SOURAV  JAIN    18      SENIOUR PROJECT MANAGER
1004    CHRIS   WAYNE   15      TEST MANAGER
1005    ANEES   FLEXWALA        19      CONFIGURATION CONTROLLER
1006    GOPAL   KRISHNA         13      SR.BIGDATA ARCHITECT
1007    RAMYA   SRI     8       BIGDATA LEAD
1008    MUNI    KRISHNA         5       SR.HADOOP DEVELOPER
1009    RAMYA   SRI     7       SR.SPARK DEVELOPER
1010    RAM     CHANDAR         16      PROJECT MANAGER
1011    AKHILA  PRIYA   3       HADOOP DEVELOPER
1012    RAHUL   DEV     6       HADOOP ADMINISTRATOR
1013    AMIT    MISRA   4       BIGDATA TESTER
1014    PUJA    YERUNKAR        11      BIGDATA TEST LEAD
1000    RAVI    VARMA   9       TECHNOLOGY LEAD
```

1001	KOUSIK	MISTRA	11	ENTRYLEVEL ARCHITECT
1002	SAIED	FAISAL	21	DELIVERY HEAD
1003	SOURAV	JAIN	18	SENIOUR PROJECT MANAGER
1004	CHRIS	WAYNE	15	TEST MANAGER
1005	ANEES	FLEXWALA	19	CONFIGURATION CONTROLLER
1006	GOPAL	KRISHNA	13	SR.BIGDATA ARCHITECT
1007	RAMYA	SRI	8	BIGDATA LEAD
1008	MUNI	KRISHNA	5	SR.HADOOP DEVELOPER
1009	RAMYA	SRI	7	SR.SPARK DEVELOPER
1010	RAM	CHANDAR	16	PROJECT MANAGER
1011	AKHILA	PRIYA	3	HADOOP DEVELOPER
1012	RAHUL	DEV	6	HADOOP ADMINISTRATOR
1013	AMIT	MISRA	4	BIGDATA TESTER
1014	PUJA	YERUNKAR	11	BIGDATA TEST LEAD

Time taken: 0.127 seconds, Fetched: 30 row(s)

TO LOAD THE DATA INTO hive_emptable(hbase_mapping table) FROM initable(Non_Mapping Table)

hive> INSERT OVERWRITE TABLE hive_emptable SELECT * FROM initable;
Query ID = gopalkrishna_20170914005226_af713c14-1e4d-45c2-9cec-2819fce28245
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1505372852042_0001, Tracking URL =
http://ubuntu:8088/proxy/application_1505372852042_0001/
Kill Command = /home/gopalkrishna/INSTALL/hadoop-2.6.0/bin/hadoop job -kill
job_1505372852042_0001
Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0
2017-09-14 00:53:24,750 Stage-0 map = 0%, reduce = 0%
2017-09-14 00:54:01,402 Stage-0 map = 100%, reduce = 0%, Cumulative CPU 8.02 sec
MapReduce Total cumulative CPU time: 8 seconds 20 msec
Ended Job = job_1505372852042_0001
MapReduce Jobs Launched:
Stage-Stage-0: Map: 1 Cumulative CPU: 8.41 sec   HDFS Read: 5653 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 410 msec
OK
Time taken: 97.971 seconds
hive>

AFTER LOADING THE DATA INTO "hive_emptable"

hive> SELECT * FROM hive_emptable;
OK

1000	RAVI	VARMA	9	TECHNOLOGY LEAD
1001	KOUSIK	MISTRA	11	ENTRYLEVEL ARCHITECT
1002	SAIED	FAISAL	21	DELIVERY HEAD
1003	SOURAV	JAIN	18	SENIOUR PROJECT MANAGER
1004	CHRIS	WAYNE	15	TEST MANAGER
1005	ANEES	FLEXWALA	19	CONFIGURATION CONTROLLER
1006	GOPAL	KRISHNA	13	SR.BIGDATA ARCHITECT
1007	RAMYA	SRI	8	BIGDATA LEAD
1008	MUNI	KRISHNA	5	SR.HADOOP DEVELOPER
1009	RAMYA	SRI	7	SR.SPARK DEVELOPER
1010	RAM	CHANDAR	16	PROJECT MANAGER
1011	AKHILA	PRIYA	3	HADOOP DEVELOPER
1012	RAHUL	DEV	6	HADOOP ADMINISTRATOR
1013	AMIT	MISRA	4	BIGDATA TESTER
1014	PUJA	YERUNKAR	11	BIGDATA TEST LEAD

Time taken: 0.297 seconds, Fetched: 15 row(s)
hive>

++
AFTER DATA LOAD IN HIVE TABLE - VERIFICATION FROM HBASE SIDE
++

 # GO TO HBASE SHELL AND TRIGGER THE scan COMMAND LIKE BELOW

hbase(main):005:0> scan 'hbasehive_emptable'

ROW	COLUMN+CELL
1000	column=empdetails:desg, timestamp=1505375640630, value=TECHNOLOGY LEAD
1000	column=empdetails:firstname, timestamp=1505375640630, value=RAVI
1000	column=empdetails:lastname, timestamp=1505375640630, value=VARMA
1000	column=empdetails:yrsofexp, timestamp=1505375640630, value=9
1001	column=empdetails:desg, timestamp=1505375640630, value=ENTRYLEVEL ARCHITECT
1001	column=empdetails:firstname, timestamp=1505375640630, value=KOUSIK
1001	column=empdetails:lastname, timestamp=1505375640630, value=MISTRA
1001	column=empdetails:yrsofexp, timestamp=1505375640630, value=11
1002	column=empdetails:desg, timestamp=1505375640630, value=DELIVERY HEAD
1002	column=empdetails:firstname, timestamp=1505375640630, value=SAIED
1002	column=empdetails:lastname, timestamp=1505375640630, value=FAISAL
1002	column=empdetails:yrsofexp, timestamp=1505375640630, value=21
1003	column=empdetails:desg, timestamp=1505375640630, value=SENIOUR PROJECT MANAGER
1003	column=empdetails:firstname, timestamp=1505375640630, value=SOURAV
1003	column=empdetails:lastname, timestamp=1505375640630, value=JAIN
1003	column=empdetails:yrsofexp, timestamp=1505375640630, value=18
1004	column=empdetails:desg, timestamp=1505375640630, value=TEST MANAGER
1004	column=empdetails:firstname, timestamp=1505375640630, value=CHRIS
1004	column=empdetails:lastname, timestamp=1505375640630, value=WAYNE
1004	column=empdetails:yrsofexp, timestamp=1505375640630, value=15
1005	column=empdetails:desg, timestamp=1505375640630, value=CONFIGURATION CONTROLLER
1005	column=empdetails:firstname, timestamp=1505375640630, value=ANEES
1005	column=empdetails:lastname, timestamp=1505375640630, value=FLEXWALA
1005	column=empdetails:yrsofexp, timestamp=1505375640630, value=19
1006	column=empdetails:desg, timestamp=1505375640630, value=SR.BIGDATA ARCHITECT
1006	column=empdetails:firstname, timestamp=1505375640630, value=GOPAL
1006	column=empdetails:lastname, timestamp=1505375640630, value=KRISHNA
1006	column=empdetails:yrsofexp, timestamp=1505375640630, value=13
1007	column=empdetails:desg, timestamp=1505375640630, value=BIGDATA LEAD
1007	column=empdetails:firstname, timestamp=1505375640630, value=RAMYA
1007	column=empdetails:lastname, timestamp=1505375640630, value=SRI
1007	column=empdetails:yrsofexp, timestamp=1505375640630, value=8
1008	column=empdetails:desg, timestamp=1505375640630, value=SR.HADOOP DEVELOPER
1008	column=empdetails:firstname, timestamp=1505375640630, value=MUNI
1008	column=empdetails:lastname, timestamp=1505375640630, value=KRISHNA
1008	column=empdetails:yrsofexp, timestamp=1505375640630, value=5
1009	column=empdetails:desg, timestamp=1505375640630, value=SR.SPARK DEVELOPER
1009	column=empdetails:firstname, timestamp=1505375640630, value=RAMYA
1009	column=empdetails:lastname, timestamp=1505375640630, value=SRI
1009	column=empdetails:yrsofexp, timestamp=1505375640630, value=7
1010	column=empdetails:desg, timestamp=1505375640630, value=PROJECT MANAGER
1010	column=empdetails:firstname, timestamp=1505375640630, value=RAM
1010	column=empdetails:lastname, timestamp=1505375640630, value=CHANDAR
1010	column=empdetails:yrsofexp, timestamp=1505375640630, value=16
1011	column=empdetails:desg, timestamp=1505375640630, value=HADOOP DEVELOPER
1011	column=empdetails:firstname, timestamp=1505375640630, value=AKHILA
1011	column=empdetails:lastname, timestamp=1505375640630, value=PRIYA
1011	column=empdetails:yrsofexp, timestamp=1505375640630, value=3
1012	column=empdetails:desg, timestamp=1505375640630, value=HADOOP ADMINISTRATOR
1012	column=empdetails:firstname, timestamp=1505375640630, value=RAHUL

1012	column=empdetails:lastname, timestamp=1505375640630, value=DEV
1012	column=empdetails:yrsofexp, timestamp=1505375640630, value=6
1013	column=empdetails:desg, timestamp=1505375640630, value=BIGDATA TESTER
1013	column=empdetails:firstname, timestamp=1505375640630, value=AMIT
1013	column=empdetails:lastname, timestamp=1505375640630, value=MISRA
1013	column=empdetails:yrsofexp, timestamp=1505375640630, value=4
1014	column=empdetails:desg, timestamp=1505375640630, value=BIGDATA TEST LEAD
1014	column=empdetails:firstname, timestamp=1505375640630, value=PUJA
1014	column=empdetails:lastname, timestamp=1505375640630, value=YERUNKAR
1014	column=empdetails:yrsofexp, timestamp=1505375640630, value=11

15 row(s) in 0.6630 seconds

hbase(main):006:0>

INSERT 1 MORE RECORD FROM HBASE SIDE => VERIFY THE SAME RECORD IS REFLECTED IN HIVE TABLE (hive_emptable) OR NOT

hbase(main):006:0> put 'hbasehive_emptable','1100','empdetails:firstname','GOPAL'
0 row(s) in 0.2460 seconds

hbase(main):007:0> put 'hbasehive_emptable','1100','empdetails:lastname','KRISHNA'
0 row(s) in 0.0320 seconds

hbase(main):008:0> put 'hbasehive_emptable','1100','empdetails:yrsofexp',13
0 row(s) in 0.0200 seconds

hbase(main):009:0> put 'hbasehive_emptable','1100','empdetails:desg','SR.HADOOP ARCHITECT'
0 row(s) in 0.0080 seconds

hbase(main):010:0>

HIVE SIDE VERIFICATION

hive> SELECT * FROM hive_emptable;
OK

1000	RAVI	VARMA	9	TECHNOLOGY LEAD
1001	KOUSIK	MISTRA	11	ENTRYLEVEL ARCHITECT
1002	SAIED	FAISAL	21	DELIVERY HEAD
1003	SOURAV	JAIN	18	SENIOUR PROJECT MANAGER
1004	CHRIS	WAYNE	15	TEST MANAGER
1005	ANEES	FLEXWALA	19	CONFIGURATION CONTROLLER
1006	GOPAL	KRISHNA	13	SR.BIGDATA ARCHITECT
1007	RAMYA	SRI	8	BIGDATA LEAD
1008	MUNI	KRISHNA	5	SR.HADOOP DEVELOPER
1009	RAMYA	SRI	7	SR.SPARK DEVELOPER
1010	RAM	CHANDAR	16	PROJECT MANAGER
1011	AKHILA	PRIYA	3	HADOOP DEVELOPER
1012	RAHUL	DEV	6	HADOOP ADMINISTRATOR
1013	AMIT	MISRA	4	BIGDATA TESTER
1014	PUJA	YERUNKAR	11	BIGDATA TEST LEAD
1100	GOPAL	KRISHNA	13	SR.HADOOP ARCHITECT

Time taken: 0.341 seconds, Fetched: 16 row(s)
hive>